

Annalen Brunner, Kathrin Steyer

Phraseologische und phraseographische Aspekte korpusgesteuerter Empirie

1 Vorbemerkung

Das Projekt *Usuelle Wortverbindungen* (UWV) ist ein korpuslinguistisches Forschungsprojekt am Institut für Deutsche Sprache in Mannheim.¹ Es untersucht und beschreibt Mehrworteinheiten auf der Basis von statistisch berechneten Kookkurrenzprofilen. Die Besonderheit der Herangehensweise liegt darin, dass die durch mathematisch-statistische Analysen gewonnenen Korpusdaten Ausgangs- und Orientierungspunkt der linguistischen Interpretation sind. Kookkurrenzen werden auf eine Weise aufbereitet, die zum einen die Musterhaftigkeit und Systematik des Miteinandervorkommens sprachlicher Einheiten linguistisch adäquat erfasst, zum anderen aber ihre „Individualität“, ihre distinktive Qualität also, nicht verschleiert. Das aus den Korpora gewonnene Wissen soll so aufbereitet werden, dass es für unterschiedliche Adressatengruppen von Nutzen ist. Neben Linguisten, die an der Struktur der deutschen Sprache – insbesondere unter einer konstruktionsellen Perspektive – interessiert sind, sind Multiplikatoren der deutschen Sprache und Deutschlerner auf mittlerem bis gehobenen Niveau wichtige Zielgruppen, denn das Wissen um typische Wortverknüpfungen ist unabdingbar für ein „kulturell angemessenes Kommunizieren“ (Stubbs 1997, 153). Im Folgenden werden wir den Nutzen eines korpusgesteuerten Ansatzes für phraseologische und phraseographische Fragestellungen anhand aktueller Forschungsergebnisse unseres Projektes darstellen.

2 Korpusgesteuerter Erkenntnisgewinn

Die Verwendung von Korpora wird auch in der Phraseologie immer mehr zum methodischen Standard. Oft nutzen Linguisten diese empi-

¹ Nähere Informationen sind auf der Projekthomepage nachzulesen (<http://www.ids-mannheim.de/kl/projekte/uvw/>).

rische Basis jedoch vor allem dazu, bereits bestehende Hypothesen zu verifizieren, oder Korpora dienen als Quelle für Belege und Beispiele. Im Gegensatz zu diesem allgemeinen korpusbasierten Ansatz (*corpus-based*) verfolgen wir den korpusgesteuerten Ansatz (*corpus-driven*) (u.a. Sinclair 1991, Tognini-Bonelli 2001). Sprachliche Strukturen werden nicht auf der Basis von A-Priori-Hypothesen, -Klassifikationen oder -Modellen erfasst, sondern anhand systematischer, empirischer Untersuchungen großer Sprachausschnitte zunächst entdeckt und dann a-posteriori interpretiert (Steyer/Lauer 2007). Unser Ziel ist es, rekurrente Sprachgebrauchsmuster zu rekonstruieren, vor allem jene, die mittels Introspektion oder durch einzelne Beispielanalysen auf Grund ihrer Alltäglichkeit oder einer wenig ausgeprägten Idiosynkrasie schwer erfassbar sind. Die Aufdeckung usueller Phänomene dieser Art ist gerade für Deutschlerner wichtig, da ihnen die Erwerbsbiografie eines Muttersprachlers und damit die intuitive Kenntnis solcher Wendungen fehlen (Hausmann 1984, 2004).

Die wichtigste Voraussetzung für ein korpusgesteuertes Vorgehen ist der Zugriff auf eine möglichst große Menge an Sprachdaten, um von einem einigermaßen repräsentativen Ausschnitt einer Sprache, in unserem Falle des Deutschen, sprechen zu können. Am Institut für Deutsche Sprache in Mannheim verfügen wir über die weltweit größte Sammlung elektronischer Korpora mit geschriebenen deutschsprachigen Texten der Gegenwart und der neueren Vergangenheit (*Deutsches Referenzkorpus DeReKo*), die über zwei Milliarden laufende Wortformen umfasst. Über das Korpusanalysewerkzeug COSMAS II sind diese Korpora auch für externe Benutzer in großen Teilen online benutz- und einsehbar.² Mit Hilfe der auch in COSMAS II integrierten statistischen Kookkurrenzanalyse (Belica 1995) kann das systematische Miteinandervorkommen von Wörtern berechnet werden, indem ermittelt wird, welche Wortformen ungewöhnlich häufig in der Umgebung des Bezugswortes zu finden sind und somit einen statistisch signifikanten Kohäsionsgrad zu ihm aufweisen. Die in so genannten Kookkurrenzprofilen zusammengefassten und hierarchisierten Clusterstrukturen geben eine Vielzahl von Information zum usuellen Ge-

² COSMAS II lässt sich derzeit auf zwei Wegen extern nutzen: zum einen über eine WWW-Schnittstelle (<https://cosmas2.ids-mannheim.de/cosmas2-web/index.jsp>); zum anderen über die Installation der Zugangssoftware (Windows-Client: <http://www.ids-mannheim.de/cosmas2/install/>).

brauch sowohl des Bezugslexems als auch der Mehrworteinheiten, die dieses Bezugslexem enthalten.³

Da die ermittelten Kookkurrenzen auf rein statistischen Berechnungen basieren, können sie der individuellen Intuition durchaus zuwiderlaufen und müssen nicht immer der traditionellen Vorstellung von echten Phrasemen als Wortverbindungen mit einem gewissen Grad an idiomatischer Qualität entsprechen. Hier meinen wir Idiomatizität als primär semantische Kategorie im Sinne von Burger (2004). Viele der auf dieser Basis zu qualifizierenden usuellen Wortverbindungen wirken transparent und unterliegen oft nicht einmal einer strukturellen Anomalie (Fleischer 1997), z.B. *sanfter Druck*, *Kapital verspielen*, *auf die Dauer*, *geistige Heimat* oder *über das normale Maß hinausgehen*. Sie sind einfach versteh- und bildbar und erscheinen auf den ersten Blick nicht als Interferenzbarriere. Diese Kandidaten weisen allein durch einen hohen statistischen Signifikanzgrad, also das auffällige Miteinandervorkommen ihrer Komponenten, eine besondere Festigkeit auf. Solche Wortverbindungen verkörpern auf Grund ihres massenhaften Gebrauchs ebenso feste Bausteine der Kommunikation wie Idiome oder Phraseme im engen Sinne. Für die Sprachproduktion sind sie sogar oft von größerer Relevanz. Ihnen kann auch eine über-summative Qualität zugeschrieben werden, die sich aber nicht immer an dem eng semantischen Kompositionalitätsprinzip orientiert, sondern in einem viel weiteren Sinne als kommunikative Idiomatizität zu interpretieren ist. Unsere Korpusanalysen haben sehr deutlich gezeigt, dass eine Mehrwortklassifikation nach dem Grad der Idiomatizität, also danach, ob es sich um eine voll-, teil- oder nicht-idiomatische Wortverbindung handelt, kein hinreichendes Kriterium ist, um den Status einer usuellen Wortverbindung angemessen zu erklären und um als alleinige Entscheidungsgrundlage für die Lemmaauswahl in einem Wörterbuch zu dienen. Die Zuschreibung von Idiomatizität ist im Grunde abhängig davon, welches Bedeutungsspektrum jeweils für die Komponenten angesetzt wird, und dies hängt wiederum vom zu

³ Die neuen Perspektiven und Spielräume linguistischer Interpretationen von Korpus- und insbesondere Kookkurrenzdaten haben wir in den vergangenen Jahren in zahlreichen Publikationen, Vorträgen und Workshops vorgestellt und diskutiert (z.B. Steyer 2000, 2002, 2003, 2004b; Steyer/Lauer 2007). Verwiesen sei auch auf die IDS-Jahrestagung 2003 *Den Nagel auf den Kopf treffen. Wortverbindungen mehr oder weniger fest*, die diesen Ansätzen eine prominente Stellung eingeräumt hat, und auf den entsprechenden Tagungsband (Steyer 2004a).

Grunde gelegten Lesartenkonzept ab. Die Wortverbindung *Kapital verspielen* etwa kann dann als nicht-idiomatisch gelten, wenn man für *Kapital* eine Lesart im Sinne von 'Potenzial, Möglichkeiten, Fähigkeiten' ansetzt und für *verspielen* eine Lesart im Sinne von 'einen erreichten Zustand wieder gefährden' bzw. 'wieder beseitigt haben'. Der Vergleich verschiedener Wörterbücher illustriert die Problematik: Im Eintrag *Kapital* des elexiko-Wörterbuchs findet sich beispielsweise neben 'Vermögen' (finanziell, d.A.) 'Werte für Produktion' und 'Kapitalisten' keine entsprechende Lesart (<http://www.elexiko.de/>). Folgerichtig müsste man die Wortverbindung dann als teilidiomatisch auffassen. GWDS verzeichnet jedoch diese Lesart, also wäre die Wortverbindung nicht-idiomatisch. Für uns ist natürlich ausschlaggebend, welche Lesartenindikatoren in den Kookkurrenzprofilen der Basiselemente, hier *Kapital* und *verspielen*, auffindbar sind. So stützen Kookkurrenzpartner wie *wichtiges, wertvolles, eigentliches, soziales, intellektuelles K.* bei *Kapital* und *Chance, Kredit, Glaubwürdigkeit, Vertrauen, Vorsprung v.* bei *verspielen* einen nicht-idiomatischen Status der Wortverbindung.

Dass nicht immer ein enges semantisches Kriterium für die Erklärung der Usualität einer Wortverbindung herangezogen werden kann, zeigen folgende Beispiele:

- (1) *zunehmend schwieriger, groß angelegt, neu aufrollen;*
- (2) *das große Zittern, in höchster Alarmbereitschaft, es fehlt an, kein Grund zur Entwarnung;*
- (3) *sich fragen lassen müssen, manchmal fragt man sich schon, immer mehr fragen sich, anders gefragt.*

Viele derartige syntagmatische Muster scheinen aus der Sicht einer Phrasen- und Satzperspektive trivial oder unvollständig. Ihre Restriktion liegt im wiederkehrenden Gebrauch genau dieser sprachlichen Formen. So besteht die Gebundenheit der Gebrauchssyntagmen bei (2) *in höchster Alarmbereitschaft* in der im Korpus fast ausschließlich vorkommenden Superlativform, bei *es fehlt an* im obligatorischen Präpositionalkomplement *an*, bei *kein Grund zur Entwarnung* im Negationsmuster. Die Gruppe (3) verdeutlicht sehr anschaulich, dass die ermittelten Gebrauchssyntagmen in genau dieser grammatischen Form spezifische kommunikative Funktionen wie 'Aufforderung' oder 'Zweifel' erfüllen.

Eine korpusgesteuerte Analyse kann für die Phraseologie und Phraseographie eine Vielzahl neuer Erkenntnisse ganz unterschiedlicher Art liefern, z.B. über „neue“ bzw. noch nicht kodifizierte usuelle Wortverbindungen, neue Lesarten und Bewertungen, obligatorische Kerne vs. fakultative Abwandlungen, typische Kontextmuster oder Beziehungen zwischen Wortverbindungen. Wir möchten hier nur einige wenige der genannten Aspekte mit aktuellem Material aus unseren Forschungen näher beleuchten.⁴

2. 1 „Neue“ bzw. noch nicht kodifizierte usuelle Wortverbindungen

Im Zuge der Analyse von Kookkurrenzprofilen stößt man neben Erwartbarem auf Kandidaten, die unbekannt erscheinen und/oder die noch nicht in Wörterbüchern/Lehrwerken verzeichnet sind. Unbekanntheit ist allerdings ein sehr relatives und individualistisches Kriterium, da auch der Linguist – so kenntnisreich er sein mag – nur ein eingeschränktes intuitives Sprachwissen haben kann. Dieses individuelle Wissen wird dem überindividuellen Ausschnitt aus dem Sprachgebrauch, den Korpora abbilden, immer unterlegen sein. Der suchende Linguist sollte also introspektiven Urteilen wie „Kenn ich nicht, kann also nicht üblich sein“ misstrauen. Wenn die statistische Analyse ein auffälliges Vorkommen einer Wortkombination ermittelt hat, geht dieses Vorkommen naturgemäß über okkasionelle oder individuelle Kommunikationsakte hinaus (abgesehen von Befunden, die durch Datenmangel oder Verzerrungen wie Eigennamen und dergleichen auch mit erhoben werden). Neben den Wortverbindungen, die dem Linguisten unbekannt scheinen, gibt es auch solche, die ihm durchaus vertraut vorkommen. Die meisten Sprecher haben keine Probleme, ihre Bedeutung in der konkreten Rezeptionssituation adäquat zu decodieren, ohne sie jedoch schon als feste Bausteine in anderen Gebrauchskontexten wieder aktualisieren und als Formulierungseinheit einsetzen zu können. Diese Wortverbindungen sind noch nicht auf Dauer in das Gedächtnis einer Sprachgemeinschaft eingegangen. In den meisten Fällen setzen sie sich quasi unter der Hand durch, ohne dass es den Sprechern wirklich bewusst ist. Die automatische Analyse

⁴ Auf die Problematik der korpusanalytisch zu ermittelnden invarianten Kerne, ihrer Varianten und Modifikationen gehen wir an dieser Stelle nicht ein. Wir verweisen hier wiederum auf die Projektpublikationen.

sehr großer Korpora, insbesondere die statistische Kookkurrenzanalyse, ermöglicht es, solche „schwebenden Wortverbindungen“ bereits in dieser Übergangsphase zu erkennen und ihre weitere Entwicklung vom Okkasionellen zum Usuellen im Sprachgebrauch zu verfolgen. Dies sei am Beispiel der Wortverbindung *Die Erde ist eine Scheibe* näher erläutert. Ein signifikanter Kookkurrenzpartner des Bezugswortes *Erde* ist das Nomen *Scheibe*, eingebettet in die satzwertige Phrase *Die Erde ist eine Scheibe*. Natürlich ist diese Wortverbindung allgemein bekannt und es ist zu vermuten, dass sie im Kontext astronomischer und historischer Diskurse als Zitat verwendet wird. Dies wäre aber noch kein hinreichender Grund, sie als gegenwärtig usuelle Wortverbindung zu interpretieren. Erst bei Durchsicht aller Korpusstellen wird deutlich, dass diese Phrase in thematisch völlig divergierenden Kontexten vorkommt. Sprecher drücken auf ironische Weise aus, dass sie etwas für genauso unwahr, abwegig, aussichtslos, unglaublich oder aber altmodisch und überholt halten wie die Aussage *Die Erde ist eine Scheibe*. Manchmal ist damit eine Kritik verbunden.

Aber dann ein Satz wie „Glauben Sie doch einfach, dass die Erde eine **Scheibe** ist und dass Frauen Auto fahren können!“ Mannheimer Morgen, 11.03.2000

Schon die mit 0:1 verlorene Partie gegen Rumänien hatte mit 22 Millionen eingeschalteten Geräten einen neuen Soccer-Rekord erzielt. Das heutige Match, das ABC vom Kabelsender ESPN übernommen hat, um eine größere Reichweite zu gewährleisten, wird neue Maßstäbe setzen. Da macht es auch nichts, daß, wie böse Zungen bemerken, die Chancen, daß die Erde eine **Scheibe** ist, größer sind als die des US-Teams. die tageszeitung, 04.07.1994

Wenn schon, dann hätte der Arme es wohl besser bei einem „normalen Bürger“ probiert, wahrscheinlich wäre die Strafe nicht so hoch ausgefallen. Hoffen wir, dass er wenigstens mit seinem Anwalt mehr Glück hat. Wenn das Gerechtigkeit ist, ist wohl die Erde doch eine **Scheibe**! St. Galler Tagblatt, 28.08.1999

2. 2 Neue Bedeutungen und Bewertungen

Die statistisch ermittelten Partner einer Wortverbindung liefern uns zudem wertvolle Informationen zu neuen Bedeutungen und pragmatischen Aspekten, die in den bisherigen Wörterbüchern noch keinen Niederschlag gefunden haben. Die Wortverbindung *Da liegt der Hund begraben* z.B. weist neben der in Duden 11 verzeichneten Lesart „[...]

das ist der entscheidende Punkt, die Ursache der Schwierigkeiten [...]“ (Duden 11, 377) in den IDS-Korpora noch eine weitere Lesart auf: *„ein Ort, wo nichts los ist“*:

Marriott, Hyatt, Hilton. In keiner Metropole dürfen die Luxus-Hotels dieser Ketten fehlen. Aber auch in kleinen Städtchen finden sich diese Monumente mondänen Lebens. So etwa im beschaulichen Stansted, einem bescheidenen Ort im Großraum London, **in dem der Hund begraben liegt** und der mit der englischen Hauptstadt so wenig zu tun hat wie Hahn im Hunsrück mit Frankfurt. Mannheimer Morgen, 13.07.2002

Auch die Wortverbindung *billiger Jakob* hat im aktuellen Gebrauch einen sehr viel weiteren Referenzbereich als z.B. in Duden 11 verzeichnet ist: *billiger Jakob* (ugs.): *Händler, bei dem die Waren besonders billig sind [...]*“ (Duden 11, 121). Unser Vorschlag einer korpusbasierten Gebrauchsbeschreibung sieht so aus: *„Mit billiger Jakob wird jemand bezeichnet, der ein Produkt oder eine Leistung zu niedrigen Preisen, oft unter Verzicht auf Anspruch und Qualität, anbietet. Billiger Jakob wird auch im Sinne eines 'Sich-unter-Wert-Verkaufens' oder eines 'nützlichen Idioten' verwendet. In den meisten Fällen ist damit eine negative Bewertung verbunden.“* Diese erweiterte Referenz auf ein allgemein menschliches Verhalten konnte u.a. durch die Analyse solcher Umgebungsmuster ermittelt werden wie *(nicht) der b. J. sein wollen; (nicht) den b. J. spielen, machen; zum b. J. werden, als b. J. auftreten, das Image eines b. J. haben.*

2. 3 Beziehungen zwischen Wortverbindungen

Die systematische Untersuchung von Kookkurrenzprofilen erlaubt es zudem, Verwandtschaften zwischen Wortverbindungen aufzudecken. Wir verwenden bewusst die eher alltagssprachliche Formulierung Verwandtschaft, um deutlich zu machen, dass es sich bei diesen Beziehungen fast nie um tradierte paradigmatische Relationen wie Synonymie oder Antonymie handelt. Oft lassen sich Gebrauchsähnlichkeiten erkennen, die allein mit Hilfe der eigenen Intuition schwerlich zu konstruieren wären. Diese in einem Beziehungsgeflecht stehenden Wortverbindungen können durchaus auch sehr divergierende Versprachlichungen aufweisen, also z.B. nicht derselben morpho-syntaktischen Klassen angehören oder unterschiedliche Lexeme umfassen. Beim Wortverbindungs-feld des Bezugswortes *blind* lässt sich auf

Grund der empirischen Befunde beispielsweise folgendes Beziehungsnetz knüpfen:

Liebe macht blind: → [NOMEN wie Hass / Wut / Gier / Eifersucht / Leidenschaft] *macht blind* → *auf einem Auge blind* → *sich blind und taub stellen* → *blind gegenüber* [der/den... NOMEN wie Realität / Tatsache / Ungerechtigkeit].

Dieses Beziehungsnetz lässt sich grob mit dem Konzept 'kritikloses Fühlen und Handeln' fassen. Gleichzeitig bringt die vergleichende Analyse fast immer die distinktiven Nuancen hervor. So referiert ...*macht blind* vor allem auf außergewöhnliche Gefühlszustände, die eine sachliche und objektive Bewertung/Handlung beeinträchtigen, *blind gegenüber...* drückt eher das Nicht-Wahrhaben-Wollen oder 'Schönreden' von objektiv notwendigen und nicht unbedingt bequemen Gegebenheiten oder Folgen aus. Gerade diese kontextuellen Feinheiten sind für einen angemessenen Sprachgebrauch von besonderer Relevanz und werden sehr oft vernachlässigt.

Das folgende Beispiel illustriert Kontextspezifika auf der konnotativen Ebene. Die verwandten Wortverbindungen *geistiger Vater*, *geistiger Brandstifter* und *geistiger Wegbereiter* bezeichnen zwar einerseits denselben Sachverhalt – 'das Einnehmen einer Vordenkerrolle' –, unterscheiden sich aber andererseits durch das jeweilige Bewertungsspektrum. *Geistiger Vater* wird im Korpus fast ausschließlich im neutralen bis positiven Sinne verwendet. Die Wortverbindung *geistiger Wegbereiter* ist ambivalenter. Sie weist ebenfalls häufig positive/neutrale Konnotationen auf (g. W. *der Wende* / *der ökologischen Bewegung*), jedoch auch negative (g. W. *des Nationalsozialismus* / *des Terrorismus*). *Geistiger Brandstifter* hingegen ist, bedingt durch das Lexem *Brandstifter*, ausschließlich negativ konnotiert und in sehr viel stärkerem Maße emphatisch. Die Teilsynonymie besteht nur zwischen den Wortverbindungen, nicht jedoch zwischen ihren nominalen Komponenten *Wegbereiter*, *Brandstifter* und *Vater*.

Eine korpusgesteuerte lexikographische Aufbereitung von Wortverbindungen kann auf Grund der eben dargestellten Komplexität nicht immer einem starren lexikographischen Beschreibungsraster folgen, z.B. in Hinblick auf die Paraphrase eines Bedeutungskerns. In den wenigsten Fällen wird der kommunikative Sinn einer Wortverbindung durch eine eng semantische Bedeutungsbeschreibung wirklich erfasst. Gerade aber dieser – durch systematische Korpusanalyse ermittelbare

– funktionale Mehrwert macht das reiche Ausdruckspotenzial von Wortverbindungen im Vergleich zum Einwortgebrauch aus. Wenn man beispielsweise nur die Grundbedeutung von *Dienst nach Vorschrift* mit 'jemand übt seinen Beruf streng nach den Vorgaben aus' umschreibt, hat man den kommunikativen Sinn und den typischen Gebrauch dieser Wortverbindung nicht erfasst. Erst durch den folgenden Zusatz wird der eigentliche Verwendungskontext deutlich: „Dabei kann sowohl gemeint sein, dass die Ausführung der Tätigkeit sehr genau geschieht, als auch, dass über die vereinbarten vertraglich geregelten Leistungen hinaus nicht mehr (zeitlich oder inhaltlich) gearbeitet wird. In der Regel wird *Dienst nach Vorschrift* als Druckmittel eingesetzt, oft anstelle eines Streiks.“ Dieser pragmatische Aspekt tritt bei der Kontextanalyse in den typischen Umgebungsmustern ganz deutlich hervor, z.B. *nur noch D. n. V., nicht einmal mehr D. n. V., ab sofort D. n. V., mit D. n. V. drohen, D. n. V. ankündigen, zum D. n. V. aufrufen* usw.

3 Lexikographische Aufbereitung – ein Anwendungsbeispiel

Der beschriebene empirische Weg korpusgesteuerter Analyse bildete die Basis für ein exemplarisch zu verstehendes lexikographisches Pilotprojekt, dessen Ergebnisse auf unserer Internet-Portalseite *Wortverbindungen online* (<http://hypermedia.ids-mannheim.de/elexiko/wv-online/>) abgerufen werden können. Hier präsentieren wir derzeit 125 Mehrwortartikel, die zwei verschiedene linguistische Beschreibungstiefen aufweisen und damit eine Entwicklung dokumentieren, die sich im Zuge der systematischen linguistischen Aufbereitung von Kookkurrenzdaten vollzogen hat.⁵

Die Erarbeitung der ausführlichen Mehrwort-Musterartikel sollte vor allem dazu dienen, die vom Projekt UWV entwickelte lexikographische Konzeption für die korpusgesteuerte Erarbeitung elektronischer Mehrwortartikel in der Praxis zu testen. Es wurden Wortverbindungen ausgewählt, die möglichst alle relevanten Mehrwortklassen, so

⁵ Wir möchten an dieser Stelle betonen, dass dieses – vor allem korpusmethodisch ausgerichtete – Pilotprojekt für eine eher rezeptive als produktive Nutzungssituation angelegt war. Das bedeutet, dass die Ergebnisse nicht in Hinblick auf den Erwerb phraseologischer Kompetenz in einer Fremdsprache unter Verwendung einer entsprechenden mehrdimensionalen Beschreibungsmatrix – wie bei Jesenšek (2006) beschrieben – aufbereitet sind. Es bleibt zu überprüfen, welche Art der Aufbereitung von Korpusinformation in welcher Präsentationsform für welche Nutzungssituation besonders geeignet ist.

wie sie die Mehrwortforschung bisher entwickelt hat, und alle in der Mehrwortartikelstruktur enthaltenen Angaben in irgendeiner Form abdecken, z.B. verschiedene Idiomatizitätsgrade, Wortverbindungen mit mehreren Lesarten, verschiedene Mehrwortklassen wie *Zwillingsformel* oder *Spruchwort*, verschiedene Phrasentypen wie *Präpositionalphrase* oder *satzwertige Wortverbindung* oder verschiedene Realisierungen kommunikativer Funktionen wie *Warnung* oder *Kritik*.

In einem nächsten Arbeitsschritt wurden die Kookkurrenzprofile der Adjektive *blind*, *geistig*, *gesund*, *normal* und *sanft* als Beschreibungsgegenstand gewählt.⁶ Die Entscheidung für adjektivische Bezugswörter ergab sich aus der Randständigkeit, die diese Wortklasse in der bisherigen Mehrwortforschung noch einnimmt. Diese liegt darin begründet, dass sich typische Partnerwörter von Adjektiven nicht oder sehr unzureichend assoziativ erschließen lassen und sich dem Blick des Linguisten daher eher verschließen als z.B. nominale oder verbale Kookkurrenzfelder. Die beschriebenen Wortverbindungen umfassen ein breites Spektrum, von zweigliedrigen phraseologischen Kollokationen wie *blinder Passagier* oder *geistige Höhenflüge* über satzwertige Wortverbindungen wie *ein blindes Huhn findet auch mal ein Korn* bis hin zu semantisch transparent erscheinenden Wortverbindungen wie *unter normalen Umständen* oder *sanfter Übergang*. Bei den nun kompakteren Gebrauchsbeschreibungen standen vor allem die kommunikativen Aspekte der Wortverbindungen im Vordergrund. Die folgenden Abbildungen verdeutlichen, wie die korpusgesteuert erarbeiteten lexikographischen Informationen in die Online-Präsentation eingeflossen sind.

Abbildung 1 zeigt den Lemmakopf unserer Mehrwortartikel. Die Möglichkeit der Kommentierung jeder Angabe ist für uns ein wichtiges Mittel, die lexikographischen Entscheidungen für den Nutzer transparent zu machen, vor allem in Hinblick auf den korpusmethodischen Hintergrund. So bedarf bereits die Auswahl des Artikelstichworts eines Kommentars, da die Entscheidung für die Lemmaform und daraus folgend für die Variations- und Modifikationsangaben eine vorgelagerte korpusgesteuerte Analyse und Reflexion erfordert. Beim Beispiel *geistige Höhenflüge* bezieht sich dieser Kommentar auf den

⁶ Diese Adjektive gehören zum Demonstrationswortschatz des *elexiko*-Wörterbuchs, der unter folgender Adresse abgerufen werden kann (http://hypermedia.ids-mannheim.de/pls/elexiko/elexiko_suche3.demows_ansehen).

Wortverbindungen online

geistige Höhenflüge Informationen Mail an UWW UWW-Projekt

Einleitung

Das Artikelkischwort wurde in der Pluralform angesetzt, da diese häufiger belegt ist als die Singularform *geistige Höhenflug*.

Drucken

Liste der bearbeiteten Wortverbindungen

Basiskomponente(n): geistig Höhenflug

Variation(en): geistiger Höhenflug

Suchanfrage für Recherche im Korpus: Agostig (=v1 &Höhenflug)

Suchanfrage

Diese Suchanfrage bietet das genaueste Ergebnis. Wenn Sie an einem möglichst großen Modifikatorenspektrum interessiert sind, sollten Sie die lexikalischen Einheiten innerhalb eines Satzes mit folgender Anfrage suchen:

Agostig hat &Höhenflug

Sie müssen dann aber damit rechnen, dass in bestimmten Fällen auch Kontextstellen angezeigt werden, die nicht zum Verwendungsspektrum dieser Wortverbindung gehören.

Schließen

Für eine dynamische Abfrage weiterer Volltextlisten in den öffentlich zugänglichen IDS-Korpora der geschriebenen Sprache empfehlen wir oben genannte Suchanfrage, die Sie in die Suchmaske der GDMAS-Webchnittstelle eingeben können.

Abbildung 1: Ausschnitt aus Mehrwortartikel *geistige Höhenflüge*

Gebrauchsbeschreibung

Blinder Alarm wird für den Sachverhalt verwendet, wenn jemand irrtümlicherweise eine drohende Gefahr signalisiert und damit entsprechende Folgeaktionen auslöst. Zumeist bezieht sich diese Wendung auf unnötige Einsätze von Feuerwehr, Polizei oder Rettungsdiensten, z.B. bei vermeintlichen Bränden, Katastrophen oder Bombendrohungen.

Beleg(e)

Blinder Alarm wird auch in einem allgemeinen Sinne für eine fälschlicherweise als gefährlich eingeschätzte Situation verwendet.

Belege

Typische Kontextmuster

blinder Alarm sein Beleg(e)

sich um blinder Alarm handeln Beleg(e)

sich als blinder Alarm erweisen Beleg(e)

sich als blinder Alarm entpuppen Beleg(e)

sich als blinder Alarm herausstellen Beleg(e)

Kommentär(e)

Die ergänzenden Verben dieser Gruppe sind an die Wortverbindung *blinder Alarm* gebunden und kommen nie oder nur selten allein mit dem Nomen *Alarm* kombiniert vor.

Schließen

blinden Alarm geben Beleg(e)

blinden Alarm auslösen Belege

Abbildung 2: Ausschnitt aus Mehrwortartikel *blinder Alarm*

Numerus. Eine weitere *Angabe Suchanfrage für Recherche im Korpus* dient dazu, dem Nutzer den Zugang zu allen Volltextbelegen einer Wortverbindung auf dem jeweils aktuellsten Stand der öffentlich zugänglichen IDS-Korpora der geschriebenen Sprache zu ermöglichen. Die Formulierung von Suchanfragen für Wortverbindungen ist keineswegs trivial, da sie nicht nur das Beherrschen der Suchanfragesprache, sondern auch ein gewisses Vorwissen in Bezug auf die Aussagekraft von Korpusanalysemethoden voraussetzt, und zudem bereits auf Hypothesenbildungen basieren muss. Daher stellen Wortverbindungsanfragen an das Korpus nicht selten eine Barriere für Nutzer dar.

Abbildung 2 verdeutlicht die Funktion der korpusgesteuerten Kontextmuster als beschreibende Stütze für die Gebrauchsbeschreibung.⁷ Interessant ist in diesem Beispiel, dass die Kontextmuster und ihre Kommentierung einen Beitrag zu einer korpusbasierten Restriktionsangabe leisten können. Wie man an der Grafik sieht, gibt es in unseren Artikeln keinen allgemeinen Belegblock, sondern spezielle – auf die jeweilige lexikographische Angabe (z.B. die Gebrauchsbeschreibung oder das einzelne Kontextmuster) zugeschnittene – Korpusbelege.

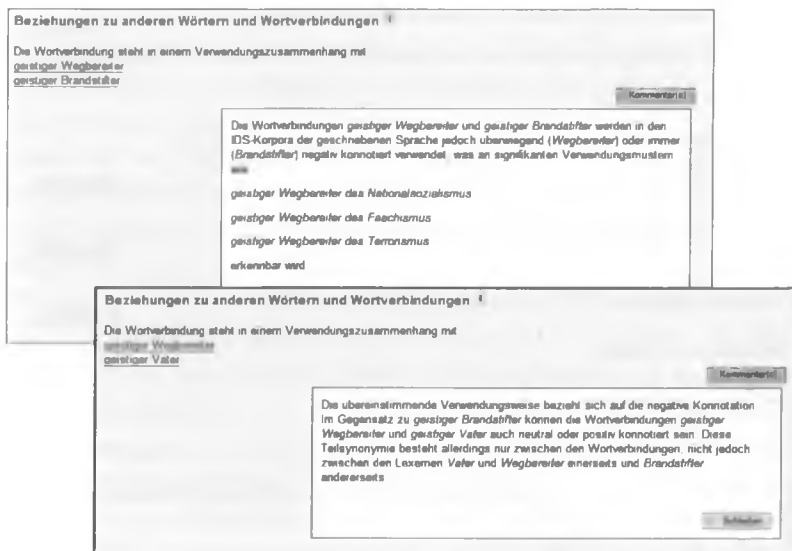


Abbildung 3: Ausschnitte aus den Mehrwortartikeln *geistiger Vater* und *geistiger Brandstifter*

⁷ Zum Konzept der Kontextmuster verweisen wir auf die entsprechenden Erläuterungen in unserem begleitenden metalexikographischen Informationstext in *Wortverbindungen online* (Steyer/Lauer 2006).

Abbildung 3 visualisiert die bereits oben beschriebene Vernetzungsangabe zu Beziehungen zwischen Wortverbindungen in den Mehrwortartikeln.

Perspektivisch ist die Aufbereitung der Daten in vernetzten Wortverbindungsfeldern geplant, in denen die sprachlichen Strukturen in ihrer gegenseitigen Bezüglichkeit adäquater dargestellt werden können. In Arbeit befinden sich das Wortverbindungsfeld *Körper*, bei dem Kookkurrenzprofile der Bezugswörter aus dem semantischen Feld der Körperteile untersucht werden sollen, sowie das Wortverbindungsfeld *Grund*, in dem ausgehend von dem Lexem *Grund* die signifikanten Partner einer weiteren Kookkurrenzanalyse unterzogen werden, um ein Feld aufzubauen, das im weitesten Sinne kausative Relationen oder entsprechende Sprachhandlungen wie *Fragen*, *Begründen*, *Zweifeln* umfasst. Diese Studien sollen einen Beitrag zur empirischen Fundierung der Konstruktionsforschung und für die linguistische Theoriebildung leisten. Des Weiteren ist langfristig die Erarbeitung eines Basisinventars usueller Wortverbindungen geplant. Hier sollen Kookkurrenzanalysen auf Basis des *Europäischen Referenzrahmens* durchgeführt, flach annotiert und für den Bereich Deutsch als Fremdsprache verfügbar gemacht werden.

4 Literatur

BELICA, CYRIL (1995): *Statistische Kollokationsanalyse und Clustering*. Korpuslinguistische Analysemethoden. Institut für Deutsche Sprache. Mannheim. <http://corpora.ids-mannheim.de>.

BURGER, HARALD (2003): *Phraseologie*. Eine Einführung am Beispiel des Deutschen. 2., überarb. Aufl. Berlin.

DUDEN 11 = *Duden*. Bd. 11 Redewendungen (2002). 2., neu bearbeitete und aktualisierte Auflage. Mannheim.

FLEISCHER, WOLFGANG (1997): *Phraseologie der deutschen Gegenwartssprache*. 2., durchges. u. erg. Aufl., Tübingen.

GWDS = *Das große Wörterbuch der deutschen Sprache* (2000). CD-ROM, basiert auf der 3., völlig neu bearbeiteten und erweiterten Auflage der Buchausgabe in 10 Bänden (1999).

HAUSMANN, FRANZ JOSEF (1984): Wortschatzlernen ist Kollokationslernen. Zum Lehren und Lernen französischer Wortverbindungen. In: *Praxis des neu-sprachlichen Unterrichts* 31, 4, 395–406.

HAUSMANN, FRANZ JOSEF (2004): Was sind eigentlich Kollokationen? In: Steyer, Kathrin (Hrsg.): *Wortverbindungen – mehr oder weniger fest*. Berlin/New York, 309–334.

JESENSEK, VIDA (2006): Phraseologie und Fremdsprachenlernen. Zur Problematik einer angemessenen phraseodidaktischen Umsetzung. In: *Linguistik online* 27, 2/06. http://www.linguistik-online.de/27_06/jerensek.html.

SINCLAIR, JOHN (1991): *Corpus, Concordance, Collocation*. Oxford.

STEYER, KATHRIN (2000): Usuelle Wortverbindungen des Deutschen. Linguistisches Konzept und lexikographische Möglichkeiten. In: *Deutsche Sprache* 28, 2, 101–125.

STEYER, KATHRIN (2002): Wenn der Schwanz mit dem Hund wedelt. Zum linguistischen Erklärungspotenzial der korpusbasierten Kookkurrenzanalyse. In: Haß-Zumkehr, Ulrike / Kallmeyer, Werner / Zifonun, Gisela (Hrsg.): *Ansichten der deutschen Sprache*. Festschrift für Gerhard Stickel zum 65. Geburtstag. Tübingen, 215–236.

STEYER, KATHRIN (2003): Korpus, Statistik, Kookkurrenz. Lässt sich Idiomatisches „berechnen“? In: Burger, Harald / Häcki Buhofer, Annelies / Gréciano, Gertrud (Hrsg.): *Flut von Texten – Vielfalt der Kulturen*. Hohengehren, 33–46.

STEYER, KATHRIN (Hrsg.) (2004a): *Wortverbindungen – mehr oder weniger fest*. Berlin, New York. [Jahrbuch des Instituts für Deutsche Sprache 2003].

STEYER, KATHRIN (2004b): Kookkurrenz. Korpusmethodik, linguistisches Modell, lexikographische Perspektiven. In: Steyer, Kathrin (Hrsg.): *Wortverbindungen – mehr oder weniger fest*. Berlin, New York, 87–116.

STEYER, KATHRIN / LAUER, MEIKE (2006): Elektronische Mehrwortartikel. Korpusgesteuerte Erarbeitung, Artikelstruktur und Online-Präsentation (lexikographische Konzeption). In: http://hypermedia.ids-mannheim.de/elexiko/wv-online/WV_info/infotext.html.

STEYER, KATHRIN / LAUER, MEIKE (2007): „Corpus-Driven“: Linguistische Interpretation von Kookkurrenzbeziehungen. In: Eichinger, Ludwig M. / Kämper, Heidrun (Hrsg.): *Sprach-Perspektiven*. Germanistische Linguistik und das Institut für Deutsche Sprache. Tübingen, 493–509.

STUBBS, MICHAEL (1997): „Eine Sprache idiomatisch sprechen“: Computer, Korpora, Kommunikative Kompetenz und Kultur. In: Mattheier, Klaus J. (Hrsg.): *Norm und Variation*. Frankfurt am Main etc., 151–167.

TOGNINI-BONELLI, ELENA (2001): *Corpus Linguistics at Work*. Amsterdam.